

# VU Research Portal

## Etnische minderheden en de Politie Intelligentie Test (PIT).

Jansen, Paul

1984

### **document version**

Publisher's PDF, also known as Version of record

### **document license**

Unspecified

[Link to publication in VU Research Portal](#)

### **citation for published version (APA)**

Jansen, P. (1984). *Etnische minderheden en de Politie Intelligentie Test (PIT)*. Rijks Psychologische Dienst.

### **General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

### **Take down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

### **E-mail address:**

[vuresearchportal.ub@vu.nl](mailto:vuresearchportal.ub@vu.nl)

P.G.W. Jansen

ETHNISCHE MINDERHEDEN  
EN DE  
POLITIE INTELLIGENTIE  
TEST  
(EEN VOORONDERZOEK)

dr. P.G.W.Jansen

maart 1984

Afd. Research  
Rijks Psychologische dienst  
Postbus 20013  
2500 EA 's-Gravenhage

ETHNISCHE MINDERHEDEN EN DE POLITIE INTELLIGENTIE TEST  
(EEN VOORONDERZOEK)

Inhoud

Samenvatting

1. Inleiding

1.1. Discriminatie door tests?

1.1.1. Predictiebias

1.1.2. Selectiebias

1.1.3. Testbias

1.2. Aanpak in het onderhavige onderzoek

2. Beschrijving en codering van de gegevens

3. Analyses

3.1. Kruistabellering

3.2. Regressieanalyse

3.3. Variantieanalyse

3.3.1. Eénweg

3.3.2. Tweeweg

3.3.3. Drieweg

4. Conclusies en discussie

Literatuur

Figuren en tabellen

## Samenvatting

In de inleiding wordt bias (een technische term voor discriminatie) van een test op zich onderscheiden van bias door een bepaald gebruik van de test. Na uitleg van twee vormen van gebruikersbias, predictiebias en selectiebias, wordt geconcludeerd dat het onderhavige onderzoek, waar geen criteriumgegevens voorhanden zijn, zich uitsluitend richt op testbias.

Op grond van gerefereerde literatuur wordt tot de volgende aanpak besloten: Nagegaan zal worden of en in welke mate zekere uit eerder onderzoek als relevant naar voren komende achtergrondvariabelen (bijvoorbeeld opleiding) een modererende invloed hebben op de testscores. Op deze wijze zouden twee vragen beantwoord kunnen worden. Ten eerste: Voldoet die subgroep allochtonen die qua achtergrond overeenkomt met de normgroep (op ethnische achtergrond na) aan de testnormen? En ten tweede: Worden de testscores door de achtergrondvariabelen beïnvloed volgens uit de literatuur bekende patronen, bijvoorbeeld stijgt de score met het nivo van de opleiding? De laatste vraag stelt in wezen de constructvaliditeit van de test voor allochtonen aan de orde.

Middels variantieanalyses kan op de laatste vraag zonder meer een bevestigend antwoord gegeven worden. Het antwoord op de eerste vraag is genuanceerder: Allochtonen met in Nederland genoten HAVO-opleiding of hoger blijken moeiteloos de testnorm te overschrijden, voor allochtonen met in Nederland genoten MAVO-opleiding geldt dat bij 7 van de 8 subtests niet ethnische achtergrond maar geslacht de testscore beslissend beïnvloedt. Slechts bij één test (analogieën) rijst het vermoeden van bias tegen ethnische achtergrond. Analyse op het nivo van de items zou daarover definitief uitsluitsel kunnen geven, wat dan ook de meest aangewezen voortzetting van dit exploratieve vooronderzoek zou zijn.



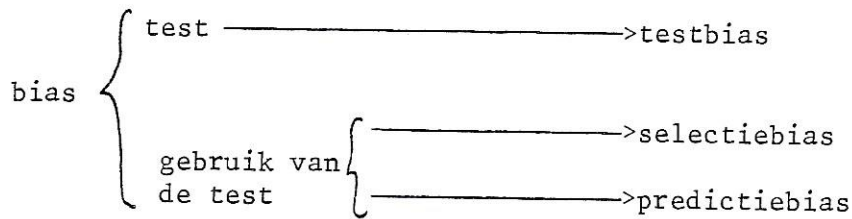
## 1. Inleiding

### 1.1. Discriminatie door tests?

De literatuur over testdiscriminatie of testbias (een technische term zonder emotionele connotatie) is uitgebreid. Zoals verderop zal blijken kenmerkt zij zich door een verfrissende exactheid (Hofstee, 1983, p. 75) hetgeen overigens niet betekent dat er geen controverses meer zouden bestaan. Aangezien de discussie 'clare et distincte' verloopt vallen die punten waar op grond van buitenwetenschappelijke criteria (bijvoorbeeld morele, ethische, politieke) keuzen gedaan moeten worden, des te beter op. Psychologisch onderzoek naar testbias is dientengevolge uiteindelijk onderzoek naar de grenzen van op puur wetenschappelijke gronden motiveerbare beslissingen in de psychologie.

In deze paragraaf wordt een, noodzakelijk beperkt, overzicht gegeven van diverse termen en definities uit de testbias-literatuur. Deze inleiding is van belang (i) omdat ze aanwijzingen heeft gegeven voor het onderhavige onderzoek, en (ii) omdat ze richting aan toekomstig onderzoek zou kunnen geven.

Allereerst dient onderscheid gemaakt te worden tussen bias van de test zelf, en bias door een bepaald gebruik van de test. Op zich hoeft een test niet te discrimineren, maar in een bepaalde context, bijvoorbeeld als predictor van een sterk cultuur-bepaalde functie, kan deze discriminerende effecten hebben. De 'gebruiksbias' van een test wordt vervolgens weer onderverdeeld in predictiebias en selectiebias. Dus in schema:



We zullen dit schema van onderen naar boven doorlopen om aldus te eindigen bij testbias. De laatste zal, bij gebrek aan verdere gegevens, centraal staan in het toegepaste deel van deze studie.

#### 1.1.1. Predictiebias

Bij predictie gaat het erom de metrische volgorde van kandidaten op een criterium Y te voorspellen op grond van de metrische volgorde van deze kandidaten op een predictor X (we beperken ons tot het geval van één predictor; generalisatie naar meerdere predictoren gaat in de hier besproken gevallen zonder moeite). Als index voor de accuratesse van de predictie wordt meestal de correlatiecoëfficiënt gebruikt, maar daarnaast ook het regressiegewicht  $b_{yx}$ , het intercept  $a_{yx}$ , en de residuele fout  $SE_{yx}$  (zie figuur 1.1). Het intercept  $a_{yx}$  geeft de initiële score op Y van iemand die minimaal scoort op X,  $b_{yx}$  kwantificeert de hoek tussen de regressielijn en de X-as (hoe groter, des te scherper de voorspelling), en  $SE_{yx}$  indiceert hoeveel de met behulp van de regressielijn  $\hat{Y} = a_{yx} + b_{yx}X$  voorspelde scores  $\hat{Y}$  nog afwijken van de werkelijke scores Y.

Er is sprake van predictiebias als een van de genoemde predictiecoëfficiënten verschillende waarden aanneemt in verschillende groepen, bijvoorbeeld in een groep van autochtonen (AUT) en een groep van allochtonen (ALL). In het vervolg zullen we steeds de groepen ALL versus AUT contrasteren.

Een bekend overzicht van onderzoek naar predictiebias is het bijna 800 pp. tellende boek van Jensen (1980). Jensen concludeert dat in het

geval van goed geconstrueerde intelligentietests,  $r_{yx}$  (de correlatie),  $b_{yx}$ ,  $a_{yx}$ , en  $SE_{yx}$  niet significant variëren over verschillende ethnische groepen. Predictiebias zou dus vrijwel niet voorkomen. Er is echter van verschillende kanten kritiek op het overzicht van Jensen geuit (vergelijk bijvoorbeeld Horn en Goldsmith, 1981), maar het voor ons belangrijkste bezwaar tegen de vergelijking van de correlatie  $r_{yx}$  tussen groepen is dat het zuivere predictiemodel in de praktijk zelden of nooit opgaat. De context van psychologisch-diagnostische advisering is vrijwel altijd selectie in plaats van predictie.

### 1.1.2. Selectiebias

Selectie verschilt van predictie doordat de scores op predictor X en criterium Y niet absoluut genomen worden, maar gerelateerd worden aan criteria van minimaal acceptabele prestatie. Zo wordt er bij het diagnostisch gebruik van de Politie Intelligentie-Test (PIT) gewerkt met afkapscores: wie lager scoort dan een van tevoren vastgestelde minimale grens voldoet niet, ook al is hij/zij de topscorer van de geteste groep. En bij latere functiebeoordelingen in het politievak wordt vanzelfsprekend impliciet of expliciet gewerkt met maatstaven voor minimaal acceptabel functioneren.

Zo bezien is de eenvoudigste vorm van selectie: één predictor X, één criterium Y, en één afkapscore ('cut-off')  $X_c$  op X en  $Y_c$  op Y.

Een verbale definitie van selectiebias is snel te geven. Volgens de Uniform Guidelines on Employee Selection Procedures (1978, p.38301) is er sprake van selectiebias als ALL lager scoort dan AUT op X, terwijl het verschil AUT-ALL op X géén verschil op Y weerspiegelt, waarbij Y een bepaalde maat van arbeidsfunctioneren vertegenwoordigt. Volgens deze definitie hoeft een gemiddeld scoreverschil van ALL en AUT



op de test op zich niet op selectiebias te duiden (wel eventueel op testbias), omdat immers deze achterstand door een passende selectieprocedure hersteld zou kunnen worden, bijvoorbeeld door een lagere afkapscore  $X_c$  te kiezen voor ALL dan voor AUT:  $X_c^{ALL} < X_c^{AUT}$ . Dit voorbeeld laat al doorschemeren dat het al dan niet discriminerend werken van een test sterk afhangt van de gebruikte selectieprocedure. Er is tot nu toe dan ook geen consensus over een standaard operationele uitwerking van selectiebias. Petersen (1980, p.110f) bespreekt vele voorstellen van eerlijke (de technische term is 'faire') selectieprocedures, zoals het regressiemodel (Cleary, 1968), het constante ratio-model (Thorndike, 1971), het conditionele waarschijnlijkheidsmodel (Cole, 1973), het gelijke waarschijnlijkhedenmodel (Linn, 1973), het gelijke risico's-model (Einhorn en Bass, 1971), en het cultuurgemodificeerd criteriummodel (Darlington, 1971).

Al deze modellen vertrekken van een rekenkundige definitie van faire selectie en geven vervolgens de daarbij passende selectieprocedure. Zo wordt bijvoorbeeld in het constante ratiomodel van Thorndike de selectie fair genoemd als het quotiënt

$$\frac{\text{kans}(X > X_c | \text{ALL})}{\text{kans}(Y > Y_c | \text{ALL})}$$

voor ALL dezelfde waarde aanneemt als voor AUT. We gaan hier niet in op de voor- en nadelen van de verschillende definities, volstaan in plaats daarvan met een verwijzing naar Petersen en Novick (1976), die aantoonde dat de verschillende definities van faire selectie onderling strijdig zijn. Dat betekent dat een selectieprocedure altijd wel discrimineert volgens een van de genoemde modellen: is aangetoond dat aan het Thorndikecriterium van constante ratio's is voldaan, dan is bijvoorbeeld het regressiemodel vrijwel zeker geschonden, en is een procedure fair in de

zin van het regressiemodel dan is hij vermoedelijk weer unfair volgens het conditionele waarschijnlijkheidsmodel.

Voordat men onderzoek doet naar de selectiebias van de PIT zal men het dus eens dienen te worden over het fairnesscriterium. De op dit moment vanuit beslissingstheoretisch oogpunt meest acceptabele selectieprocedure is het volgens de principes van de decisietheorie maximaliseren van de verwachte utiliteit van het selectieproces (Gross en Su, 1975). Een in dit geval relevant concreet geval van zo'n procedure is snel geschetst.

Stel men berekent in de groepen ALL en AUT de regressielijnen  $Y = a_{yx} + b_{yx} X$ . Dan is de selectieprocedure vanuit beslissingstheoretisch oogpunt fair als de afkapgrenzen  $X_c^{ALL}$  en  $X_c^{AUT}$  zo gekozen worden dat

$$a_{yx}^{ALL} + b_{yx}^{ALL} X_c^{ALL} = Y_c^{ALL},$$

$$a_{yx}^{AUT} + b_{yx}^{AUT} X_c^{AUT} = Y_c^{AUT}.$$

In woorden: Van iemand die in een van beide groepen precies op de afkapgrens scoort wordt verwacht dat deze op het criterium de voor die groep geldende minimaal nog acceptabele score zal behalen. Kiest men voor de criteriumafkapgrenzen in beide groepen bepaalde waarden (bijvoorbeeld dezelfde:  $Y_c^{ALL} = Y_c^{AUT}$ ), dan kunnen  $X_c^{ALL}$  en  $X_c^{AUT}$  berekend worden door substitutie in bovenstaande vergelijkingen. Aldus zijn de test- en criteriumafkapgrenzen zo gekozen dat altijd aan fairness in de zin van de decisietheorie is voldaan.

Echter, de consequentie van deze keuze van een definitie van selectiefairness zou wellicht kunnen zijn dat blijkt dat  $a_{yx}^{ALL} \neq a_{yx}^{AUT}$ . Dat betekent dat de gekozen selectieprocedure fair is volgens de beslissingstheoretische definitie, maar unfair (of biased) volgens het uitgangspunt van het predictiemodel.

Bovendien is het ook goed mogelijk dat blijkt dat om aan beslissingstheoretische fairness te voldoen de testafkapgrenzen zo gekozen moeten worden dat  $X_C^{ALL} < X_C^{AUT}$ , hetgeen vanuit een andere visie op fairness bezien weer als oneerlijk betiteld kan worden.

Dit voorbeeld illustreert dat bij elke keuze van een selectie-model vraagtekens op grond van een of andere definitie van fairness gezet kunnen worden, en dat derhalve een test vanuit het ene gezichtspunt fair kan zijn, maar vanuit het andere niet. Voor welke definitie van eerlijkheid, rechtvaardigheid, non-discriminatie, fairness, unbiasedness men kiest kan alleen maar de uitkomst zijn van een buiten- (of, zo men wil, boven-) wetenschappelijke discussie tussen diverse belanghebbenden, het is kortom een beleidskeuze. De eventueel unfaire gevolgen van zo'n beleidsmatig gemotiveerde selectiebeslissing dienen bij het maken van die keuze al meegewogen te worden, en, vooral, de negatieve maar onvermijdelijke neveneffecten van zo'n keuze voor één specifiek selectiemodel dienen expliciet door het beleid tot zijn verantwoordelijkheid gerekend te worden.

### 1.1.3. Testbias

We verlaten nu de context van de selectie en beschouwen de test als een op zichzelf staand meetinstrument. De meetpretentie van de test behelst dat de manifeste score  $X$  indicatief is voor een latent vaardigheidsnivo  $V$ . De relatie  $X$ - $V$  wordt gespecificeerd (en gecontroleerd) middels een meetmodel, waarbij dat van de klassieke testtheorie tot nu toe nog het meest gangbare is. Van belang voor het volgende is dat  $V$  per definitie het ware, echte nivo representeert, terwijl  $X$  de manifest gemeten waarde is op grond waarvan  $V$  wordt afgeleid. Eventuele bias zit dan in een verstoorde relatie tussen  $V$  en  $X$ : Iemand met hoge  $V$  sjoert desondanks een lage  $X$ , en krijgt dus, ten onrechte, een lage  $V$  toe-



gemeten.

Met behulp van deze begrippen is een definitie van testbias snel te geven (en deze is, i.t.t. selectiebias, algemeen aanvaard, cf. Petersen, 1980, p.108f): Een test is vrij van bias (is 'zuiver' of 'cultuurvrij') als de relatie X-V in zowel ALL als AUT dezelfde is. Hoe voor de hand liggend deze definitie ook mag zijn, hij bevat een evidente cirkel: Hoe de ware score V vast te stellen als dat per definitie alléén kan langs de omweg van de manifeste indicator X van V die wel eens een onzuivere ('<sup>is</sup>biased') weergave van V zou kunnen zijn?

Bij onderzoek van testbias gaat het derhalve erom het echte vaardigheidsnivo vast te stellen los van de vermoede imperfectie van de geobserveerde testscore. Er zijn verschillende methoden geopperd om V te benaderen (Rudner et al., 1980a,b; Scheuneman, 1980; Mellenbergh, 1982), echter de ideale procedure voor onderzoek van testbias is Raschanalyse (Jansen, 1983), vergelijk de voorbeelden in Scheuneman (o.c., p.142). Al deze methoden onderzoeken testbias via eventuele itembias: De globale testscore X wordt voor waar aangenomen, en de bij die X behorende ware score V wordt berekend. Vervolgens worden twee groepen, ALL versus AUT, met dezelfde V gevormd, en gaat men na hoe die groepen scoren op een gegeven item i. Als de itemscores  $X_i^{ALL}$  en  $X_i^{AUT}$  significant verschillen, wordt de relatie  $V-X_i$  ( $X_i$  is meestal een proportie) beïnvloed door de ethnische achtergrond, hetgeen op bias van het betreffende item wijst: Latent even vaardige personen scoren verschillend op hetzelfde item.

Helaas kon in het onderhavige onderzoek testbias niet als itembias onderzocht worden. Het aantal personen van de ALL-groep was op zich groot genoeg (98) om itemkenmerken in ALL betrouwbaar vast te stellen (zelfs na opsplitsing naar geschatte V), maar de individuele itemscores waren niet op korte termijn beschikbaar. Aangezien voor de AUT-groep (d.w.z. de normgroep van de PIT) de itemkenmerken bekend zijn, is onderzoek van itembias van de PIT-subtests theoretisch goed uitvoerbaar. Het



is ook een voor de hand liggend vervolg op dit onderzoek. De meest aangewezen methode om dit itemonderzoek op korte termijn te kunnen voltooiën is de benaderingsmethode van Scheuneman (voorbeeldig uitgelegd in Scheuneman, 1980, pp.145-148), welke uitermate makkelijk is toe te passen.

Zoals opgemerkt waren in het onderhavige onderzoek echter alleen de totale testresultaten beschikbaar. In de volgende paragraaf wordt uiteengezet hoe nochtans is getracht greep te krijgen op eventuele bias van de PIT.

## 1.2. Aanpak in het onderhavige onderzoek

Een interessant voorbeeld van onderzoek naar predictiebias is de studie van Van der Flier en Drenth (1980). Ze onderzochten de invloed van een aantal biografische variabelen, zoals ethnische achtergrond, geslacht, kwaliteit van de lagere school, linguïstische achtergrond en beroep van de vader, op de correlationele relaties tussen enerzijds scores op een tiental nonverbale intelligentietests (X) en anderzijds proefwerkresultaten (Y). Het bleek dat de belangrijkste moderator van de predictieve validiteiten  $r_{xy}$  niet de ethnische achtergrond was, maar het beroep van de vader, het geslacht, en de kwaliteit van de lagere school.

Op grond van soortgelijk onderzoek concludeert Novick (1980, p.124f) dat men in plaats van ethnische achtergrond als criterium voor eventuele bias de term 'kansarmen' ("the disadvantaged") zou moeten gebruiken. Ethnische achtergrond is immers een onveranderbare biografische factor, terwijl kansarm zijn ontleed kan worden in kenmerken als te lage scholing of gebrekkige beheersing van de taal, waarop wel tot actie zou kunnen worden overgegaan, bijvoorbeeld in de vorm van bijscholing.

De definitie van een variabele A die iemand tot kansarme bestempeld is: Ten gevolge van A is het nivo van de persoon in kwestie niet het maximaal haalbare nivo. Merk op dat in deze benadering de testprestatie op zich niet meer wordt aangevochten (die is objectief laag), maar de interpretatie ervan in het licht van de kandidaat zijn/haar achtergrond (de vaardigheid kan eigenlijk beter, er zou meer in zitten, etc.). De consequentie van deze aanpak is evenwel dat personen met dezelfde biografische achtergrond ook ongeveer dezelfde testcores moeten behalen. Of in het geval van het onderhavige onderzoek: Die personen uit ALL die wat betreft biografische achtergrond niet of nauwelijks verschillen van de corresponderende achtergrond van de groep (AUT) waarop de normen van de PIT zijn gebaseerd, dienen ongeveer 'normaal' (d.w.z. volgens de normen) te scoren. Als dat zo is, is ethnische achtergrond een irrelevante variabele geworden, en kan met meer recht gekeken worden naar factoren die tot relatief lage testcores aanleiding geven, zoals bijvoorbeeld een vooropleiding die veel lager is dan het gemiddelde nivo van de PIT-normgroep.

Daarmee is meteen de strategie van het onderhavige onderzoek geschetst. Op grond van de gerefereerde en andere literatuur is een selectie gemaakt van die biografische variabelen die naar verwachting modererend op de testscore zouden kunnen werkem. Vervolgens is nagegaan of dat inderdaad het geval is, en met name of die subgroep uit ALL die qua relevante biografische kenmerken niet verschilt van de PIT-normgroep, gemiddelde ongeveer even hoog scoort als de normgroep. Het onderzoek is dus gericht op het opsporen van achtergrondvariabelen die de testscore mede (naast de echte vaardigheid V) bepalen. Op deze wijze zou de globale variabele 'ethnische achtergrond' ontleed kunnen worden in aantoonbaar relevante factoren, en zouden bovendien aanwijzingen gevonden kunnen worden wat te doen met personen die

laag scoren op de PIT (bijvoorbeeld bijscholen in de functie o.i.d.).

Deze benadering maakt tevens duidelijk in hoeverre variatie van de PIT-scores met een variabele als opleiding overeenkomt met de uit de testliteratuur bekende verbanden: de testscore stijgt in het algemeen met de opleiding. In wezen wordt in deze studie derhalve de constructvaliditeit van de PIT voor de ALL-groep onderzocht. De centrale vraag is: Verschilt de PIT voor ALL en AUT voor wat betreft zijn constructvaliditeit?

## 2. Beschrijving en codering van de gegevens\*

Van 98 allochtonen (ALL) werden de volgende gegevens verzameld: de eindscores (klassescores met in de normgroep standaardgemiddelde 50 en standaard-standaarddeviatie 15) op de 8 PIT-subtests (WIT: woorden invultest, SOM: sommentest, STIL: stilleestest, OBJV: objectvergelijken, ANAL: analogieën, CAUS: causale reeksen, PLR: plaatjesreeksen, TMI: technisch-mechanisch inzicht), de totale PIT-eindscore (EIND), en een 6-tal biografische variabelen: geslacht, leeftijd, ethnische achtergrond, verblijfsduur in Nederland, opleiding in het buitenland of in Nederland, opleidingsnivo. Sommige biografische variabelen werden gehercodeerd tot een 3- of 4-categoriële variabele opdat zich in elke categorie een voor statistisch betrouwbare conclusies voldoende aantal personen zou bevinden. Een overzicht van de diverse indelingen en eventuele ordinale coderingen vindt men in tabel 2./

De categorieën van de acht<sup>v</sup>grondvariabele opleiding bevatten de volgende schooltypes:

LTS: lbo, lts-b, lts-c, lhno, leao,

MAVO: ulo, mulo, mulo-b, mavo-3, mavo-4, meao, mts,

---

\*. Met dank aan A. Dutman, die voor de verzameling vande gegevens zorgdroeg.



HAVO: havo, lyceum, pedagogische academie, atheneum.

LTS staat dus voor een lage opleiding, MAVO voor een middelbare, en HAVO voor een relatief hoge opleiding. Nominaal gezien zou MAVO moeten overeenkomen met het opleidingsnivo van de PIT-normgroep.

De in de 3e kolom van tabel 2.1 gegeven aantallen tellen per variabele vanwege soms ontbrekende scores niet altijd op tot 98. In de in dit rapport besproken analyses varieert het effectieve aantal personen tussen de 90 en 95. In de laatste kolom van de tabel zijn de scores vermeld welke aan de diverse categorieën van een variabele werden toegekend. Dat kan natuurlijk alleen bij achtergrondvariabelen met een natuurlijke ordening zoals leeftijd en opleiding. De codering is dan ook alleen gebruikt voor de in paragraaf 3 te bespreken regressieanalyse, welke zuiver exploratief van aard was. De wezenlijke bevindingen van dit onderzoek zijn gebaseerd op variantieanalyses, waarin de achtergrondvariabelen uitsluitend nominaal worden opgevat.

### 3. Analyses

In paragrafen 3.1 en 3.2 worden de relaties tussen de biografische variabelen onderling onder de loep genomen. Aldus worden de variantieanalyses van paragraaf 3.3 voorbereid, waarin de relaties tussen de uit de paragrafen 3.1 en 3.2 als belangrijk naar voren komende achtergrondvariabelen en de PIT worden bestudeerd.

#### 3.1. Kruistabellering

Studie van alle bivariante kruistabellen van de achtergrondvariabelen leert: De ethnische groep 'overige' is voor het merendeel reeds zeer lang (meer dan 10 jaar) in Nederland. Dientengevolge heeft deze

groep ALL voor het overgrote deel de opleiding in Nederland ontvangen. Nochtans verschilt het nominale nivo van de opleiding van 'overige' niet van dat van de opleidingen van Surinamers en Antillianen: De 3 ethnische groepen zijn in gelijke mate verdeeld over de 3 opleidingsnivo's. Opleiding en plaats van opleiding blijken stochastisch onafhankelijke variabelen (chikwadraat is 2.1 bij 4 vrijheidsgraden, zie tabel 3.1). Het is bijvoorbeeld niet zo dat in het buitenland genoten opleidingen vaker van het nominale LTS-nivo zijn. Blijkbaar is zowel in het buitenland als in Nederland de verdeling van de bevolking over de 3 onderscheiden nivo's van opleiding dezelfde.

De bevinding dat opleiding en plaats van opleiding onafhankelijk van elkaar variëren over de ALL-groep is van belang omdat het impliceert dat hun eventuele invloed op de PIT-score zich naar verwachting onafhankelijk van elkaar zal doen gelden.

### 3.2. Regressieanalyse

De diagnostisch gezien belangrijkste score van de PIT is EIND. In deze exploratieve fase van het onderzoek is derhalve alleen een regressieanalyse van EIND op in aanmerking komende achtergrondvariabelen uitgevoerd. Als mogelijke predictoren van de eindscore werden gekozen geslacht, plaats van opleiding, en opleiding. Verblijfsduur is weggelaten omdat deze voor een groot deel blijkt samen te vallen met plaats van opleiding (correlatie is .65); de laatste variabele wordt, ook vanwege de bevindingen van de kruistabellering, als meer fundamenteel gezien.

Regressieanalyse van EIND op alleen plaats van opleiding en opleiding levert een multiple correlatie van .48. Blijkens de analyse leveren opleidingsplaats en opleidingsnivo een vrijwel even grote bijdrage aan EIND (5.1 versus 6.5 als gewichten). Toevoeging van geslacht doet de

multiple correlatie licht stijgen tot .55.

Analyses van de overige PIT-scores zijn hier achterwege gelaten; in de volgende paragraaf komen die wel aan de orde.

### 3.3. Variantieanalyse

Uit de literatuur blijken opleiding en geslacht naar voren te komen als belangrijke potentiële beïnvloeders van de testscore. Bovendien blijkt volgens de kruis- en regressieanalyse dat opleiding, plaats van opleiding en geslacht de belangrijkste PIT-score voor een groot deel bepalen. Tenslotte geldt nog de overweging dat bij de constructie van de PIT-normen niet gelet is op ethnische achtergrond, noch op verblijfsduur, noch op geslacht, maar wel (expliciet) op opleiding en (impliciet) op plaats van opleiding.

Uit deze drie bevindingen volgt dat naar verwachting prestaties op de PIT vooral bepaald zullen worden door opleiding, plaats van opleiding, en, in mindere mate, geslacht. Dit vermoeden is getoetst middels reeksen variantieanalyses met als afhankelijke variabele steeds één van de 9 PIT-scores en als onafhankelijke variabelen de biografische kenmerken plaats van opleiding, opleiding, en geslacht. Onderzoek van eventuele effecten van ethnische achtergrond heeft geen zin omdat de vraagstelling zich richt op prestatieverschillen tussen ALL en AUT, en niet op die tussen ALL-groepen onderling. Overigens zijn er wel verschillen binnen ALL ('overige' scoort over het algemeen hoger dan Surinamers of Antillianen), die zijn echter steeds terug te voeren op verschillende nivo's op de achtergrondvariabele plaats van opleiding (vergelijk paragraaf 3.1).

Bij 3 onafhankelijke variabelen is maximaal een 3-weg variantieanalyse mogelijk. De bespreking schrijdt stapsgewijs in die richting.



### 3.3.1. Eénweg

Tabel 3.2 geeft een overzicht van de resultaten. Als significantiegrens is de conventionele waarde .05 gebruikt. Ter vergelijking zijn in de onderste rij de gemiddelden ( $\bar{X}$ ) en standaarddeviaties ( $s_x$ ) van ALL op de PIT-subtests en op EIND vermeld. Die onderste rij maakt duidelijk dat in alle gevallen ALL significant lager scoort dan de normgroep (toetsing middels: is  $\bar{X} + 2s_x / \sqrt{n}$  groter dan het normgemiddelde 50?). Vooralsnog is de enige conclusie die aan die resultaten verbonden kan worden dat ALL geen toevalstrekking uit AUT vertegenwoordigt.

Tabel 3.2 maakt duidelijk dat eigenlijk alle tests wel beïnvloed worden door één of meer achtergrondvariabelen. De eindscore is dientengevolge de resultante van zowel opleiding, als plaats van opleiding, als geslacht. Alle verbanden zijn in de op grond van de literatuur verwachte richting ( $H > M > L$ ,  $MA > V$ ,  $N > B$ ), en wat dat betreft vertonen de PIT-scores van ALL géén afwijkend patroon ten opzichte van AUT. In beide gevallen geldt dezelfde vorm van constructvaliditeit.

De tabel bevestigt dat opleidingsplaats en -nivo de meeste invloed op de PIT-prestaties hebben. Voor beide achtergrondvariabelen geldt dat de effecten niet geconcentreerd zijn op een van de PIT-factoren semantisch (WIT, STIL, ANAL, CAUS), figuraal (OBJV, PLR, TMI), en symbolisch (SOM): De effecten zijn willekeurig verdeeld over de bij elke intelligentiefactor behorende tests.

### 3.3.2. Tweeweg

In die gevallen waar twee of meer achtergrondvariabelen de PIT-scores significant bepalen moeten twee- en driewegvariantieanalyses uitwijzen of en in welke zin er interactie is tussen de variabelen. Zo zou het wellicht het geval kunnen zijn dat een combinatie van opleiding in het buitenland én LTS-opleiding meer is, in negatieve zin,



dan de optelsom van de negatieve effecten van beide biografische karakteristieken apart.

Er viel maar in één geval interactie te bespeuren, en wel bij de gecombineerde effecten van opleidingsplaats en -nivo op ANAL. Aan het einde van deze paragraaf wordt dit geval kort besproken. In alle andere gevallen beïnvloeden de achtergrondvariabelen onafhankelijk van elkaar de PIT-scores. Zo levert het gevolgd hebben van een LTS-opleiding zowel voor mannen als voor vrouwen en zowel voor hen die deze opleiding in het buitenland gevolgd hebben als voor hen die deze in Nederland volgden, eenzelfde 'handicap' op voor wat betreft de relatieve PIT-scores (relatief ten opzichte van MAVO en HAVO). Hoeveel deze handicap ongeveer bedraagt kan men vinden in tabel 3.3. Een voorbeeld ter illustratie.

Blijkens de onderste rij van tabel 3.3 is het totaalgemiddelde van ALL op STIL ongeveer 35. Derhalve zal iemand met een opleiding in het buitenland (-6) en met een LTS-opleiding (-8) naar verwachting ongeveer 21 scoren ( $35 - 6 - 8 = 21$ ). Het verschil tussen de verwachte gemiddelde score 21 en het geobserveerde gemiddelde 26.7 in de cel (opleiding in buitenland, LTS) ontstaat door steekproeffout (de standaardfout van het laatste gemiddelde is 5.6); interactie is immers afwezig in dit geval. Daarentegen zal iemand met een HAVO-opleiding in Nederland naar verwachting 50 scoren, hetgeen precies gelijk is aan de norm (het geobserveerde gemiddelde is 54.6). Bemerkt het verschil tussen dit gedifferentieerde paar van scores (21, 50) en het globale STIL-gemiddelde 35.

Dergelijke berekeningen zijn voor alle PIT-subtests te maken, met dien verstande dat bij ANAL rekening gehouden moet worden met interactie (zie verderop), en dat bij CAUS en EIND mogelijk 3-weg interactie voorkomt (overigens zeer onwaarschijnlijk).

Tabel 3.3 wijst uit dat het effect van de achtergrondvariabelen per test varieert; zo is de negatieve invloed van vrouw-zijn op de SOM-score 2x zo groot dan op de EIND-score. Globaal gesproken geldt echter dat de beide relevante variabelen opleiding en plaats van opleiding (met name relevant omdat middels bijscholing de eventuele negatieve effecten ervan weggenomen zouden kunnen worden) in termen van klassescores de prestaties op de PIT met gemiddelde 14 punten beïnvloeden in negatieve zin (LTS in het buitenland), en met 11 punten in positieve zin (HAVO in Nederland). Of nog globaler: een optimale opleiding verhoogt het ALL-gemiddelde met ongeveer één PIT-standaarddeviatie, een minimale verlaagt de ALL-score op de PIT met één standaarddeviatie.

Opleidingsnivo en -plaats hebben dus een aanzienlijk effect op de PIT-prestaties. Om dit nog eens te illustreren is tabel 3.4 opgenomen, waarin de geobserveerde gemiddelden op EIND in de diverse opleidingscondities staan vermeld. Merk op dat de randgemiddelden in deze tabel vrijwel gelijk zijn aan de met behulp van tabel 3.3 berekende randgemiddelden (bijvoorbeeld: HAVO, EIND: verwacht randgemiddelde is  $33+6=39$ , geobserveerd is 39.5). De celgemiddelden wijken door steekproeffouten af van de berekende gemiddelden.

Tenslotte: Bij ANAL was er sprake van sterke interactie in de tweeweganalyse. Nadere analyse wijst uit dat die op rekening van de HAVO-groep komt. In het buitenland genoten scoort HAVO het laagst op ANAL (LTS: 25.1, MAVO: 25.6, HAVO: 19.3), in het buitenland én Nederland genoten stijgt de relatieve HAVO-positie op ANAL naar de tweede plaats (LTS: 18.0, MAVO: 44.6, HAVO: 26.0), en uitsluitend in Nederland genoten verovert HAVO veruit de eerste plaats op ANAL (LTS: 26.9, MAVO: 36.5, HAVO: 50.0). Vermoedelijk zijn de geringe aantallen B+N (3 voor LTS, 5 voor MAVO, en 3 voor HAVO) debet aan dit typische interactie-

effect.

### 3.3.3. Drieweg

Blijkens de tweeweganalyses beïnvloeden zowel opleidingsplaats als opleidingsnivo, als geslacht het nivo op de PIT. Alleen in het geval van CAUS en EIND zijn alle drie de bijdragen significant, dus alleen in deze twee gevallen zou een 3-weg variantieanalyse zin hebben. Zo'n 3-weg variantieanalyse verdeelt de totale groep (+95) over 18 cellen, zodat gemiddeld 5 personen per opleidingsplaatsXopleidingsnivoXgeslacht-combinatie verwacht mogen worden. Dat is te weinig voor betrouwbare statistische conclusies, en derhalve is afgezien van een 3-weg variantieanalyse.

Niets verhindert ons echter op descriptieve basis de 3-weg tabellen met 18 cellen te bestuderen; daartoe beperken we ons in deze paragraaf. Leidraad voor de analyse van de 3-weg tabellen is de in paragraaf 1.2 gestelde vraag naar de PIT-scores van die subgroep uit ALL die nominaal aan de PIT-normgroep equivalent is: Opleiding in Nederland genoten, en minimaal MAVO-nivo. De PIT-gemiddelden van de relevante subgroep uit ALL zijn vermeld in tabel 3.5. Studie van deze tabel leidt tot de voorzichtige, want soms op geringe aantallen gebaseerde conclusies:

- ALL die qua opleiding overeenkomen met de normgroep (MAVO in Nederland) scoren op 6 van de 8 tests (en derhalve ook wat betreft eindscore) gemiddeld te laag op de PIT;
- wordt onderscheid gemaakt naar geslacht dan blijkt dat mannen met MAVO-Nederland op 3 van de 8 tests (en op EIND) te laag scoren, en vrouwen met MAVO-Nederland op 4 van de 8 (en op EIND);
- in de totale groep MAVO-Nederland liggen alle ALL-gemiddelden statistisch rond het normgemiddelde;



-wordt echter onderscheid gemaakt naar geslacht dan blijken vrouwen met HAVO-Nederland op 3 van de 8 tests (en op EIND) onder het normgemiddelde te zitten; deze conclusie is evenwel gebaseerd op slechts 2 personen!

In de slotparagraaf worden deze bevindingen nader uitgewerkt.

#### 4. Conclusies en discussie

De analyse van de driewegtabel 3.5 wijst uit dat personen met HAVO-Nederland als achtergrond de PIT zonder problemen aankunnen. In drie gevallen scoort de vrouwelijke subgroep te laag, maar aangezien het hier slechts om 2 personen gaat kunnen er nauwelijks conclusies aan dit resultaat verbonden worden. Van bias van de PIT ten opzichte van ALL is hier in ieder geval geen sprake.

De groep met MAVO-Nederland scoort als geheel gemiddeld te laag op de PIT, vergeleken met de normgroep. Wordt echter een uitsplitsing gemaakt naar geslacht, dan blijkt dat -alweer gemiddeld- mannen onder de norm scoren op tests die uitsluitend laden op de verbale factor (WIT, ANAL, CAUS), en dus ook op EIND, en dat vrouwen beneden het normgemiddelde presteren op de rekenkundige (SOM) en de figurale tests (PLR, TMI), en derhalve ook op EIND. Slechts op één test (ANAL) zitten beide groepen te laag. Dit alles wijst erop dat het te laag scoren op de PIT van de MAVO-groep voor 7 van de 8 subtests niet het gevolg van de ethnische achtergrond kan zijn (want daarin verschillen mannen en vrouwen in dit onderzoek niet), maar mogelijk toegeschreven moet worden aan geslachtsverschillen in intelligentie. De bevinding dat mannen beter scoren dan vrouwen op rekenkundige en figurale tests en vrouwen beter dan mannen op verbale tests correspondeert met soortgelijke resultaten van eerder onderzoek. "Sex differences in spatial tests and certain mathematical reasoning tests have been well documented" concluderen

Tenopyr en Oeltjen (1982, p.594) in hun recente overzicht van de stand van zaken op het gebied van de selectiepsychologie, en ze vervolgen: "Thus, it appears that women are at a disadvantage competing for jobs demanding mathematical and spatial abilities". Merk op dat het hier niet om sex-bias hoeft te gaan, maar mogelijk om reële verschillen (wellicht cultuur-bepaald) in de aard van intellectueel functioneren.

Op ANAL en op EIND scoren zowel mannen als vrouwen onder het gemiddelde. Het verschil tussen de eindscore 39.6 in de groep MAVO-Nederland en de eindscore 50 in de normgroep lijkt op zich niet zo groot, maar in termen van verwachte percentages afgewezen kandidaten is het aanzienlijk. Als bijvoorbeeld een afkapgrens van 45 op EIND wordt gehanteerd, dan zal in de normgroep, waar de scores verdeeld zijn volgens  $N(50,15)$ , 37% afvallen. In de groep ALL met MAVO-Nederland zal, onder de aanname van  $N(39.6,13.5)$  65% afvallen, dus bijna twee keer zo veel. Nochtans kan bij deze gegevens de te lage waarde van EIND niet toegeschreven worden aan de ethnische achtergrond van de kandidaten. De te lage waarde van EIND in de totale groep is het onvermijdelijke gevolg van de lage score op een aantal PIT-subtests. Eerder is evenwel inzichtelijk gemaakt dat de beneden-norm prestaties op PIT-subtests veeleer toegeschreven moeten worden aan de modererende invloed van geslacht op 7 van de 8 PIT-subtests dan aan ethnische achtergrond.

Uiteindelijk blijft ANAL over. Op deze test zit de MAVO-groep, ook na opsplitsing voor wat betreft opleiding in Nederland en geslacht, ver beneden het normgemiddelde. Bij vervolgonderzoek naar itembias zal dan ook speciale aandacht naar ANAL moeten uitgaan.

Tot besluit nog de opmerking dat het ons inziens noodzakelijk is informatie te verzamelen over de ALL-posities op achtergrondvariabelen

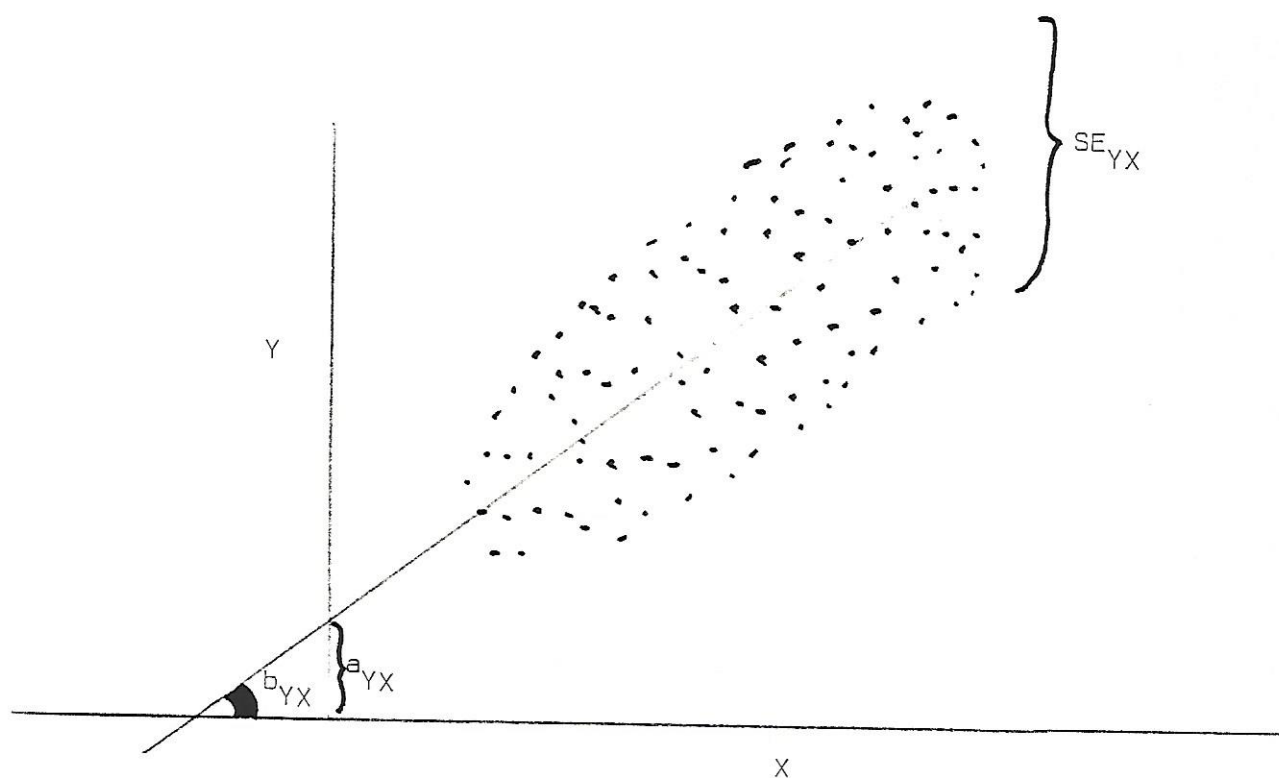
die in dit onderzoek ontbraken, maar waarvan het op grond van eerder onderzoek (bijvoorbeeld de in paragraaf 1.2 besproken studie van Van der Flier en Drenth) aannemelijk is dat ze een veel grotere invloed hebben op de prestaties op een test voor algemene intelligentie dan ethnische achtergrond. Terecht wijzen leden van minderheidsgroepen erop dat het laatste kenmerk voor wat betreft selectie op basis van psychologische tests oninteressant dient te zijn, en wellicht dat deze studie althans die indruk achterlaat: Dat onderzoek naar bias wezenlijk erop is gericht ethnische achtergrond in de context van selectie-tests terug te brengen tot wat het zou moeten zijn: een irrelevant, nietszeggend kenmerk.

Literatuur

- Cleary, T.A. Test bias: Prediction of grades of negro and white students in integrated colleges. Journal of Educational Measurement, 1968, 5, 115-124.
- Cole, N.S. Bias in selection. Journal of Educational Measurement, 1973, 10, 237-255.
- Darlington, R.B. Another look at "culture fairness". Journal of Educational Measurement, 1971, 8, 71-82.
- Einhorn, H.J., & Bass, A.R. Methodological considerations relevant to discrimination in employment testing. Psychological Bulletin, 1971, 75, 261-269.
- Equal Employment Opportunity Commission, Civil Service Commission, Department of Labor, and Department of Justice. Adoption by four agencies of Uniform Guidelines on Employee Selection Procedures, Federal Register, 1978, 43, 38290-38315.
- Flier, H., van der, & Drenth, P.J.D. Fair selection and comparability of test scores. In: Van der Kamp, Langerak, & De Gruijter, 1980, 85-101.
- Gross, A.L., & Su, W.H. Defining a "fair" or "unbiased" selection model: A question of utilities. Journal of Applied Psychology, 1975, 60, 345-351.
- Hofstee, W.K.B. Selectie. Utrecht: Het Spectrum, 1983.
- Horn, J., & Goldsmith, H. Reader be cautious: A review of 'Bias in mental testing'. American Journal of Education, 1981, 89, 305-329.
- Jansen, P.G.W. Rasch analysis of attitudinal data (Akademisch Proefschrift). Den Haag: Rijks Psychologische Dienst, 1983.
- Jensen, A.R. Bias in mental testing. London: Methuen, 1980.
- Kamp, L.J.Th., van der, Langerak, W.F., & Gruijter, D.N.M., de. Psychometrics for educational debates. New York: Wiley, 1980.



- Linn, R.L. Fair test use in selection. Review of Educational Research, 1973, 43, 139-161.
- Mellenbergh, G.J. Contingency table models for assessing item bias. Journal of Educational Statistics, 1982, 7, 105-118.
- Novick, M.R. Policy issues of fairness in testing. In: Van der Kamp et al., 1980, 123-137.
- Petersen, N.S. Bias in the selection rule -Bias in the test. In: Van der Kamp et al., 1980, 103-122.
- Petersen, N.S., & Novick, M.R. An evaluation of some models for culture-fair selection. Journal of Educational Measurement, 1976, 13, 3-29.
- Rudner, L.M., Getson, R.R., & Knight, D.L. Biased item detection techniques. Journal of Educational Statistics, 1980, 5, 213-233. (a)
- Rudner, L.M., Getson, R.R., & Knight, D.L. A Monte Carlo comparison of seven biased item detection techniques. Journal of Educational Measurement, 1980, 17, 1-11. (b)
- Scheuneman, J. Latent-trait theory and item bias. In: Van der Kamp et al., 1980, 145-151.
- Tenopyr, M.L., & Oeltjen, P.D. Personnel selection and classification. Annual Review of Psychology, 1982, 33, 581-618.
- Thorndike, R.L. Concepts of culture-fairness. Journal of Educational Measurement, 1971, 8, 63-70.



Figurr 1.1. Lineaire predictie van het criterium Y door de predictor X.  
De punten vertegenwoordigen de geobserveerde paren  $(X,Y)$ , de  
lijn de bestpassende lineaire voorspelling  $\hat{Y} = a_{YX} + b_{YX}X$  van  
Y per nivo van X.

Tabel 2.1. Coderingen/categoriseringen van enige achtergrondvariabelen

achtergrond- variabele	categorieën	aantal	eventuele codes
leeftijd	< 20 jaar 20-22 jaar 23-26 jaar ≥ 27 jaar	20 34 17 22	
ethnische achtergrond	Suriname Antillen Molukken, Indië, Turkije Marokko (overige)	62 17 14	
verblijfsduur	< 4 jaar 4-10 jaar > 10 jaar	29 32 32	
plaats van opleiding	buitenland (B) buitenland + Nederland (B+N) Nederland (N)	36 12 46	1 2 3
opleiding	LTS (L) MAVO (M) HAVO (H)	30 47 17	1 2 3
geslacht	vrouw (V) man (MA)	27 63	

Tabel 3.1. Relatie tussen de biografische variabelen opleiding en plaats van opleiding  
(chikwadraat is 2.1 bij 4 vrijheidsgraden)

		opleiding			
		LT S	MAVO	HAVO	ΣM
plaats van opleiding	buitenland	9	21	6	36
	buitenland + Nederland	4	5	3	12
	Nederland	17	21	8	46
	ΣM	30	47	17	94

Tabel 3.2. Resultaten van éénweg-variantieanalyses. Significantienivo is .05, een blanco cel impliceert géén significante verschillen tussen de biografische groepen. Voor de afkortingen, zie tabel 2.1.

	WIT	SM	STL	OBJV	ANAL	CAUS	PLR	TM	EIND
plaats van opleiding			N>B	N>B	N>B	N>B	N>B B+N>B	N>B B+N>B	N>B B+N>B
opleiding	H>L M>L **	H>L M>L	H>M>L		H>L M>L	H>L M>L	H>L		H>L M>L
geslacht		MA>V				MA>V		MA>V	MA>V
gemiddelde	33.7*	34.9*	35.0*	38.6*	30.7*	30.1*	30.5*	31.7*	33.4*
standaarddeviatie	17.4	17.4	17.9	18.0	15.2	15.1	17.5	17.7	13.0

\*. Gemiddelde is significant kleiner dan het normgemiddelde 50.

\*\*. Dit betekent dat HAVO en MAVO niet wezenlijk verschillen qua WIT-score, maar dat de gemiddelde WIT-scores in beide opleidingsgroepen significant groter waren dan de gemiddelde WIT-score in de LTS-groep.

Tabel 3.3. Afgeronde effectgroottes van drie achtergrondvariabelen op de afgeronde PIT-gemiddelden (bij ANAL dient rekening gehouden te worden met interactieeffecten). Effectgroottes zijn alleen vermeld voor die combinaties waar blijkens de variantieanalyse significante invloed van de achtergrondvariabele op de PIT-score optrad.

	WIT	SOM	STIL	OBJV	ANAL	CAUS	PLR	TMI	EIND
opl. B			-6	-7	-6	-5	-9	-6	-6
v. B+N			-3	+1	+2	+3	+4	+5	+2
opl. N			+5	+6	+5	+3	+6	+3	+4
opl. L	-8	-7	-8		-5	-8	-4		-5
M	+3	+1	+1		+2	+3	-0		+1
H	+3	+8	+10		+4	+6	+7		+6
gesl. V		-10				-5		-9	-5
MA		+4				+2		+4	+2
afgerond totaal- gemid- delde *	34	35	35	38	31	31	31	32	33

\*. Zodanig afgerond dat de herberekende (afgeronde) randgemiddelden dicht bij de werkelijke (afgeronde) randgemiddelden liggen.

Bijv. : HAVO, STIL: berekend gemiddelde is  $35+10=45$ , geobserveerd gemiddelde is 45.4.



Tabel 3.4. Gemiddelden (\*), standaarddeviaties (\*\*), en groepsgroottes (tussen haakjes) van EI ND onder diverse opleidingscondities.

		plaats van opleiding			
opleiding		B	B+N	N	totaal
	L	25.2*	28.0	29.4	27.9
		7.6**	8.7	9.7	8.9
		(9)	(3)	(17)	(29)
	M	28.2	40.0	39.6	34.6
		10.0	7.9	13.5	13.5
		(21)	(5)	(21)	(47)
	H	30.3	35.3	47.9	39.5
		14.9	8.5	12.1	14.6
		(6)	(3)	(8)	(17)
	tot.	27.8	35.4	37.3	33.4
		10.2	13.4	13.5	13.0
		(36)	(11)	(44)	(93)



Tabel 3.5. Gemiddelde PIT-scores voor MAVO, HAVO binnen Nederland, opgesplitst naar geslacht.

	opleiding in Nederland					
PIT	MAVO			HAVO		
	totaal	man	vrouw	totaal	man	vrouw
WIT	37.3*	36.2*	39.0	44.1	41.2	53.0
SOM	38.8*	45.5	27.8*	48.9	53.8	34.0
STIL	43.3	44.5	41.4	54.6	54.2	56.0
OBJV	47.2	43.4	53.4	50.9	50.5	52.0
ANAL	36.5*	37.7*	34.5*	50.0	54.0	38.0*
CAUS	38.4*	39.2*	37.1	44.1	48.8	30.0*
PLR	38.6*	40.2	36.0*	40.6	44.7	28.5
TMI	37.3*	40.9	31.5*	38.1	46.5	13.0*
EIND	39.6*	40.8*	37.6*	47.9	51.2	38.0*
aantal personen	21	13	8	8	6	2

\*. Gemiddelde is significant kleiner dan 50.